

**ASSESSING COST-EFFECTIVENESS OF COMPUTER-BASED
TECHNOLOGY IN PUBLIC ELEMENTARY AND SECONDARY SCHOOLS**

by

David Stern and Guy Cox
School of Education
University of California, Berkeley

January 8, 1987

This document was prepared under a contract with the Office of Technology Assessment, Congress of the United States, for the assessment, *Power On: New Tools for Teaching and Learning*. The conclusions are those of the authors. The document does not necessarily reflect the analytical findings of OTA, the Advisory Panel, or the Technology Assessment Board.

5952-23

Assessing Cost-Effectiveness of Computer-Based
Technology in Public Elementary and Secondary Schools

Statement of Problem

There are always plenty of proposals for improving education, but never enough money to implement most of them. Recommendations for raising teachers' salaries, reducing class size, reorganizing curriculum, and promoting new educational technologies all compete for the limited budgets available.

As OTA embarks on its comprehensive assessment of computer-based technology in elementary and secondary schools, a highly pertinent debate has flared up among researchers, on the issue of relative cost-effectiveness. Specifically, Levin, Glass and Meister have published evidence comparing the cost-effectiveness of reducing class size, lengthening the school day, cross-age tutoring, and computer-assisted instruction. A full account of the study appeared in two 1984 project reports from Stanford University. The results did not become known to a wide audience until June, 1986, when Phi Delta Kappan published a summary of the study. The Kappan also took the rather unusual step of publishing a critique along with the study itself.

The Levin-Glass-Meister evidence is important for the OTA study not only because it indicates cross-age tutoring can be more cost-effective than computer-assisted instruction for raising achievement in the elementary grades, but also because it is the first published study that has actually made systematic cost-effectiveness comparisons of several distinct educational programs. Such studies have not been done in the past because reports on effects of educational treatments have very rarely included data on costs. With a few exceptions, researchers have not kept track of the resources required to deliver different educational programs in elementary and secondary schools. In approaching its current

assignment to study computer-based technology in education, therefore, OTA is obliged to consider a kind of analysis that has seldom been done before.

Analyzing cost-effectiveness in education will require attention to three major issues:

- Educational outcomes are many-dimensional. For instance, students usually receive instruction in several different subjects each day. This complicates the problem of comparing interventions which are specific to one subject (such as tutoring, or a particular program of computer-assisted instruction) against interventions that promote instruction in any or all subjects (e.g., smaller class size or a longer school day).
- Educational treatments also are many-dimensional. For instance, two programs of computer-assisted instruction may have the same stated objective, but differ in potentially important ways (such as whether students work individually or in groups, with the regular teacher or a special instructor, etc.). This complicates the problem of defining a set of programs for which to compute average "effect size."
- Computer-based technology is changing fast. Both hardware and software become obsolete in a relatively short time. This has implications for measuring cost-effectiveness from actual program data.

This paper elaborates on these issues. It begins with an explanation of why cost-effectiveness studies are important: since schools lack incentives to economize, they do not routinely monitor their own efficiency or seek ways to improve it. The basic economic paradigm for assessing cost-effectiveness is presented next, followed by discussion of how the basic procedure may be extended to handle the problems of multiple outcomes, multi-dimensional treatments, and limitations of existing data. Implications for the OTA study will be presented at the end.

Efficiency and Incentives in Public Schools

Economic analysis of education addresses two main questions. First, does the monetary value of benefits produced by expenditures on education equal or exceed the value of those expenditures? Second, are schools and other educational efforts producing as much learning as possible, given their current budgets? The first question demands a benefit-cost (or cost-benefit) analysis. Examples of benefit-cost analysis in education will be given below. Most of the attention in this paper will focus on the second question, which underlies cost-effectiveness analysis.

If education were a commodity like breakfast cereal or a service like hair styling, economists would not ask these questions. Benefit-cost analysis would not be necessary because the operation of markets tends to equate the cost of producing a box of cereal or a haircut with its value to the consumer, measured by his or her willingness to pay. In other words, consumers do their own benefit-cost analysis, and this leads to efficient allocation of resources since members of society are not seriously affected by each others' choice of breakfast cereal or hair style.

However, education is obviously different from these private goods and services, since members of society are seriously affected by each others' education. Educating one person benefits other members of society because education improves that person's competence as a citizen, parent, soldier, and neighbor. Adam Smith himself urged government support of local schools for this reason. Education also enables some individuals to make cultural and scientific contributions, for which the material compensation they receive is less than the benefit to society as a whole. Alfred Marshall saw this as a reason for public support of schools. Most contemporary economists still agree that education produces non-marketable benefits, and therefore warrants some degree of public funding because people would buy too little of it otherwise. (For contemporary discussions, see Benson, 1978; Cohn, 1979; Haveman and Wolfe, 1984.) For these

reasons, nations everywhere in the world today not only support education with public funds, but even require school attendance by law.

Compulsory schooling laws are intended to give people less choice about education than they have about haircuts or breakfast food. In addition, using government-operated schools as the principal mechanism for channeling tax revenues into education further restricts choice. At the elementary and secondary levels in the United States, the only way for households to obtain the maximum subsidy for schooling is to send their children to public schools. Although private schools are exempt from paying taxes, for the most part their expenses are not paid by government. If some sort of voucher scheme made it possible for private schools to be fully supported by public revenues, more households probably would choose private schools. At present, only about ten percent do so.

Given that public schools face little competition at the elementary and secondary level, economists are not surprised to learn that these schools generally do not appear to operate in a cost-effective manner. More than a hundred studies have produced empirical evidence bearing on the cost-effectiveness of public schools. The nature of these studies will be described below. Hanushek (1986), surveying this evidence, concluded:

"If we think of schools as maximizing student achievement, the preceding evidence indicates that schools are economically inefficient, because they pay for attributes that are not systematically related to achievement." (p. 1166)

This "does not come as a great surprise" to Hanushek or other economists (see Levin, 1976) because public schools are not like business firms that lose revenues and eventually go out of business if they fail in market competition. Public school revenues come from local, state, and federal sources. These payments are not conditioned on any measure of what students learn. Local elections concerned with school taxes may sometimes be influenced by perceptions of school quality, but this is a highly imperfect mechanism. Furthermore, public expenditures for education tend to be driven by perceived needs, so that states and the federal government tend to add money when failures become more

evident. In addition to programs serving special needs, states also distribute a large amount of money simply on the basis of student headcount. This creates some incentive to maintain attendance, but not to increase learning.

This line of thinking has led some economists and others to argue for education vouchers, so that a school would receive less revenue if fewer families chose to attend. A less sweeping kind of proposal, also intended to make public schools more accountable, is to declare a school educationally bankrupt when it grossly fails to meet its obligations to students (see Committee for Economic Development, 1985). A bankrupt public school would then be taken over and run by the state. Recently, the National Governors Association endorsed a proposal like this. The lack of incentive for public schools to operate cost-effectively is evident to many people and elected officials.

Also evident is the lack of material incentive for individual teachers and administrators to produce as much student learning as possible. Students have to be active collaborators in the process, of course, and motivating students who are required to be there is part of the challenge for teachers and administrators. But how well they surmount this challenge and find ways for students to learn does not determine how much the teachers or administrators are paid. This evident lack of material incentive to do the job as well as possible has prompted many experiments with "merit pay" for teachers (Hatry and Greiner, 1985; Cohen and Murnane, 1985). Currently, career ladders and lattices for teachers are being designed all around the country.

What has destroyed or distorted many merit pay plans in the past is the difficulty of defining and measuring the behavior that makes some teachers more effective than others. Simply relying on supervisors' ratings leads to charges of unfair subjectivity, but developing other kinds of evidence (dossiers, observations and review by committees) tends to become inordinately time-consuming. Recently some districts have tried awarding bonus pay for teachers in groups rather than as individuals. Group incentives are intended to avoid competition between individuals, and foster collaboration instead. Another idea

receiving much current attention is to create a national board that would certify teachers, for both basic and advanced levels of competence (Carnegie Forum on Education and the Economy, 1986). The prototype examinations now being developed for this proposed national board presumably will embody some conception of what good teachers do.

For the time being, teachers and administrators lack clear external standards of good practice, either as individual members of a profession or as employees in a bureaucratic organization. Teachers and administrators who work hard do so because they are intrinsically motivated, and for this parents of public school children ought to be grateful.

Whether it is possible to create extrinsic incentives for school professionals that reinforce their intrinsic motivation, and do not lead merely to teaching the test or other counterproductive games, remains an open question (see Stern, 1986). But the fact remains that decision-makers in public schools have rarely been given any material inducement to maximize students' learning. This has important implications for OTA's current inquiry into uses of computers in schools.

One implication of not explicitly seeking to maximize students' learning is the lack of systematic innovation by schools themselves. Schools are constantly innovating, but they generally do not keep track of the results. Innovations therefore tend not to produce cumulative understanding as fast as might happen if innovation were coordinated and monitored more systematically. In particular, schools almost never examine the relative cost-effectiveness of different techniques for achieving given objectives. Empirical studies of cost-effectiveness in education have been done mainly by researchers outside the schools themselves. Some of these have dealt specifically with use of computers, but in order to get definitive answers about the cost-effectiveness of computers for achieving various educational purposes, additional research will be necessary. The following sections describe the research on which further studies can build.

Another implication of schools' lack of systematic incentive to maximize students' learning is that results of cost-effectiveness studies by outside researchers or agencies will

not necessarily influence what schools do. Change itself is costly and time-consuming. What seems to make sense to an outside analyst does not necessarily persuade insiders to act. As we analyze cost-effectiveness, therefore, we should try to see it from the point of view of each participant in the educational process.

Basic Paradigm for Analyzing Cost-Effectiveness

Economists analyzing cost-effectiveness in education begin with a basic theory of the "firm." A firm is a single decision-making entity which uses one or more inputs, or "factors of production," to produce one or more outputs. Inputs include such things as the time of various kinds of employees, for which the firm must make explicit payments. Inputs also include equipment or facilities which the firm may own but for which it incurs an "opportunity cost," since physical capital could be liquidated and the money invested in other ways. Outputs include goods or services for sale.

Profit is the difference between the revenues obtained from sale of outputs and the cost of producing those outputs. If a firm is maximizing profit, input quantities must be adjusted such that the value of output produced by the last unit of each input is equal to the unit price of that input. (For a standard derivation of this result, see Nicholson, 1983.) For instance, if a dentists' office could increase its revenues by \$30 as a result of hiring a hygienist for one additional hour, and if the cost to the office for that additional hour were only \$15, then the office could increase profit by increasing hygienists' time. If the office added more and more hygienists' time, eventually space limitations and other constraints would reduce the amount of additional revenues from an additional hour of hygienists' time to \$15, which equals the hygienists' hourly wage. At that point, if nothing else changed, the office would be maximizing its profit from hygienists.

A firm is said to be economically inefficient if it does not maximize profit, given available technology. There are two distinct ways in which a firm may be inefficient. First, it may fail to obtain the maximum feasible output from the inputs it is using. This is

called technical inefficiency (see Levin, 1976; Jamison and Lau, 1982). Simply wasting time or materials, due to ignorance or carelessness, creates technical inefficiency.

A firm may be technically efficient but still fail to maximize profit, if it buys the wrong combination of inputs or sells the wrong amounts of output. This is called allocative inefficiency. Technical inefficiency can be observed in the physical process of production itself, without knowing the prices of inputs and outputs. In contrast, measuring allocative inefficiency requires information about what various inputs cost and what price can be obtained for output. In other words, technical efficiency is the absence of physical waste; allocative efficiency requires a correct response to prices. Both are necessary for profit maximization.

This basic paradigm for analyzing efficiency can be applied to public schools, even though we cannot compute a school's "profit" when students do not pay tuition and there is no explicit price for students' learning. It is possible to compare schools on the basis of either technical or allocative efficiency.

Educational production functions. Most of the research by economists has focused not on schools' technical efficiency, but on their allocative efficiency. To test whether schools are efficient in allocating their budgets, economists have compared the cost of various inputs with their estimated productivity. Productivity is usually estimated from a multiple regression in which measures of students' background and prior achievement, along with school inputs, predict students' current achievement. The research on "educational production functions" has been reviewed by Lau (1979), Benson (1978), Cohn (1979), and Hanushek (1986), among others.

Coefficients on school inputs in an estimated production function can be compared directly with the cost to schools of purchasing these inputs. For example, Levin (1970) used a production function (estimated by Hanushek, 1968) in which teachers' score on a test of verbal ability and teachers' years of experience were included, along with other school inputs and characteristics of students, as predictors of achievement in sixth grade.

Levin also estimated the additional salary cost associated with higher verbal scores and years of teaching experience. (Standard salary schedules for teachers award higher pay automatically as experience increases through the first two decades or so. Verbal or other test scores are not explicit factors on salary schedules, but the higher cost of recruiting teachers with higher test scores was estimated by Levin and has also been substantiated by other studies, e.g., Manski, 1985.) Comparing the cost of these two characteristics of teachers with their estimated effect on students' achievement, Levin discovered that the annual cost of producing a one-point gain in a sixth-grade student's achievement was \$26 (in 1968 dollars) when the gain was obtained by employing teachers with higher verbal scores, while producing the same gain by employing teachers with more experience would cost five to ten times as much. This indicates that schools could produce higher student achievement -- with a given budget -- by recruiting teachers with high verbal scores instead of more experience.

Hanushek's (1986) recent review of more than a hundred production function studies came to the general conclusion, quoted above, that public schools are generally inefficient because they pay for inputs which do not seem to produce achievement by students. Specifically, effects of the following inputs on students' achievement have not usually been found to be positive and statistically significant: higher educational attainment by teachers (which leads to higher pay on the standard salary schedule), longer teaching experience, smaller class size, higher teachers' salary, or larger expenditures per pupil.

There are several well-known reasons why "educational production functions" should not be taken at face value. These reasons have all been carefully stated by contributors to, or critics of, this research in its early stages, during the aftermath of the 1966 Coleman report (see U.S. Office of Education, 1970), and repeated by reviewers already cited. The well-known limitations include difficulty of imputing causality in correlational studies (e.g., do more experienced teachers somehow teach better, or are they simply in a better position to choose the students they teach?); inadequacies of certain available data (e.g., the

survey analyzed in Coleman (1966) did not measure change or match individual students to individual teachers); and difficulty of analyzing more than one output at a time (see discussion below). In addition, if schools are technically inefficient (as economists believe they are), then observed correlations between inputs and outputs are probably smaller than the correlations would be if schools did not waste any resources. Pervasive technical inefficiency, therefore, implies that standard "educational production functions" do not provide a basis for accurate judgements about the degree of allocative efficiency.

Still, the results of "educational production function" studies are not as unambiguously negative as Hanushek's conclusion suggests. For instance, his table (p. 1161) shows 60 studies had findings about teachers' salary and student achievement, of which only ten are said to be statistically significant. Of these ten, nine studies showed a positive association. Among the other 50, 15 were positive, 11 negative, and 24 had unknown sign. If the "true" association between salary and student achievement were zero, how likely would this pattern of results be? Applying more formal meta-analytic procedures (see Hedges and Olkin, 1985) might not support Hanushek's conclusion that teacher salary makes no difference.

On the effect of class size, another variable which makes no difference according to Hanushek, formal meta-analysis by Glass (1984) did reach the opposite conclusion. Using more than a hundred comparisons produced by 14 experiments in which students were randomly assigned to classes of different sizes, Glass fitted a logarithmic relationship between class size and achievement. He found that cutting class size in half (from 20 to 10, 30 to 15, 10 to 5, or whatever) on average increases achievement by 0.2 standard deviations, or approximately two grade-equivalent months in elementary school.

The purpose of estimating educational production functions is to predict the change in output consequent upon a change in one input at a time. Multiple regression is the statistical procedure most commonly used. If Hanushek or anyone else wanted to do a formal meta-analysis of this research, he or she would have to solve a technical problem:

namely, the fact that the coefficient on one predictor (e.g., teacher salary or class size) depends on what other predictors are in the equation. Since different studies use different combinations of predictors, averaging estimated coefficients across studies should take this into account. Some progress in developing meta-analytic procedures for multiple regressions is reported by Ash (1985).

Experimental studies. Estimation of educational production functions has relied on naturally occurring variation in inputs and outputs among schools. As mentioned, causal interpretation of such correlational studies is often problematic. A more direct approach to estimating cost-effectiveness of school programs, which yields clearer causal inferences, is to conduct actual experiments. Educational researchers have reported on thousands of experimental or quasi-experimental studies -- of which there are now hundreds dealing with computers -- but only a few of these contain information on costs.

One set of experimental studies that kept careful account of costs was reported by Jamison, Fletcher, Suppes, and Atkinson (1976). Their report dealt with computer-assisted instruction (CAI) as a method of compensatory education for disadvantaged children. Achievement of students receiving CAI was compared with similar students who received no CAI. (The report implies that, in two of the experiments, comparison groups did receive compensatory education in some other form.) CAI was found to produce statistically significant gains. Moreover, the per-pupil cost of CAI was found to be well within the per-pupil budget available for compensatory education. The authors concluded, therefore, that CAI was feasible and (if comparison groups were receiving conventional compensatory education services) cost-effective as a form of compensatory education. They also computed how big an increase in class size would be required if compensatory funds were not available and CAI had to be financed by reducing the number of teachers. They conclude that CAI would be cost-effective as a partial replacement of teachers. However, this part of their analysis assumed class size does not affect achievement -- a debatable assumption, given the analysis by Glass (1984) described above.

More recently, Levin, Glass, and Meister (1984; see also Levin and Meister, 1986, Levin, Glass, and Meister, 1986) have combined correlational, experimental, and quasi-experimental findings in order to compare the cost-effectiveness of four different educational interventions: reducing class size, lengthening the school day, introducing CAI, and instituting a cross-age tutoring program. To estimate effects of reducing class size, the authors relied on Glass's (1984) meta-analysis of experimental studies, described above. Effects of a longer school day were estimated from results obtained by the Beginning Teacher Evaluation Study (Denham and Lieberman, 1980), in which teachers' and students' use of time was found to be correlated with students' gains in achievement. The estimated effect of CAI came from a carefully conducted experiment in Los Angeles, using drill-and-practice software produced by the Computer Curriculum Corporation (Ragosta, Holland, and Jamison, 1982). To estimate effects of a cross-age tutoring program, Levin et al. used quasi-experimental evaluation results from an exemplary program in Boise. They summarized their results in the following table (Levin and Meister, 1986, p. 748).

Estimated Effectiveness of Four Educational Interventions
in Months of Additional Achievement Gain per Year
of Instruction for Each \$100 per Student

		<u>Mathematics</u>	<u>Reading</u>
CAI		1.0	1.9
Cross-age tutoring			
Peer component		4.6	2.2
Adult component		0.8	0.5
Increasing instructional time		0.5	1.2
Reducing class size			
From	To		
35	30	1.4	0.7
30	25	1.2	0.6
25	20	1.0	0.5
35	20	1.1	0.6

This shows CAI is cost-effective at the elementary level compared to increasing instructional time, but reducing class size appears to be more cost-effective than CAI in mathematics, and the student-tutor component of the cross-age tutoring program seems to be the most cost-effective of the four innovations, in both mathematics and reading.

Niemiec, Blackwell, and Walberg (1986) criticized the study by Levin et al. for estimating the effects of CAI and tutoring based on single cases, instead of synthesizing estimates from the dozens of cases which have been published. In response, Levin et al. (1986) argued that their purpose was to compare effect size with the cost of a well-defined program. Previously published studies of CAI and student tutoring have not included information on cost. Therefore, they maintained it was more appropriate to use data on estimated effects only from the single cases for which they had obtained cost data.

The kind of analysis Levin et al. have done yields direct and obvious implications for policy and practice. Replicating this analysis with data from other cases would be highly desirable. The method could also be extended to account more explicitly for use of students' time, and to deal with multiple outcomes (see discussion below).

Benefit-cost analysis. As explained above, cost-effectiveness studies are concerned with whether schools produce the maximum amount of student learning, given limited budgets. In contrast, benefit-cost analysis asks whether the monetary value of benefits from education exceeds the amount of money spent on it. This is a question of external efficiency -- should expenditure on education increase or decrease -- as opposed to the question of internal efficiency addressed by cost-effectiveness studies -- how should the given budget be spent.

The monetary value of benefits from education has been measured mainly by increased earnings of individuals who have completed more years of schooling. Education is seen as an investment in "human capital" (Becker, 1975). The cost of producing human capital includes teachers' salaries and other direct costs of schooling. The cost also

includes students' time, which has an opportunity cost especially for older students, because they could be earning money if they were not in school.

By comparing the cost of the investment with the additional earnings obtained as a result of additional schooling, it is possible to compute the rate of return on the investment. This has been done for various levels of schooling, at different times and in various places (see Psacharopoulos, 1985). The social rate of return counts public expenditures on schooling as part of the cost, and includes additional tax payments by more highly educated individuals as part of the benefit to society. In less developed countries, real social rates of return exceed 25 percent at the primary level; at the secondary and post-secondary levels the rates of return are around 15 percent. In the industrialized countries, real social rates of return average around ten percent at the secondary and postsecondary levels.

Benefit-cost analysis can also be used to evaluate particular educational programs. For example, Weisbrod (1965) evaluated a high school dropout-prevention program; Hu, Lee, and Stromsdorfer (1971) analyzed the additional costs and benefits of high school vocational education; Garms (1971) evaluated a program (Upward Bound) for encouraging low-income high school students to attend college; Mallar (1982) and others have analyzed Job Corps training; Psacharopoulos and Loxley (1985) and Psacharopoulos (1986) have compared vocational versus general academic curricula at the secondary and post-secondary levels. All these studies have used earnings as the measure of benefit from additional schooling.

The use of earnings to measure benefit from schooling is based on human capital theory, which assumes that schooling enables individuals to earn more money because it actually renders them more productive. Critics have questioned this assumption, arguing that schools may merely sort and screen individuals, awarding degrees and diplomas only to those who were most able at the outset. Arrow (1973) and Spence (1973) showed it is conceivable that more able individuals find it profitable to attend school even if school does nothing to make them more productive; in this extreme case, the true social return to

schooling would be negative. However, more recent theoretical models suggest that certain kinds of screening or credentialing mechanisms may actually induce too little investment in schooling rather than too much (Weiss, 1985). Empirical attempts to measure the screening effect have indicated that it accounts for some, but probably less than half, of the private return to schooling (Cohn, 1979, p. 46). It is difficult to test directly whether schooling increases individuals' productivity, because in most industrialized work settings the interdependence of jobs makes it impossible to measure one individual's output separately. The clearest evidence that education does increase productivity has come from studies of individual farmers in less developed countries (Jamison and Lau, 1982).

To the extent that schools merely screen rather than teach, the additional earnings associated with further schooling overstate the social benefit from education. On the other hand, education also produces a number of benefits other than earnings. These are not easily translated into monetary values, and are therefore not included in quantitative benefit-cost calculations. Nonetheless, they are real. Some of these non-monetary benefits from education accrue mainly to the people who get the schooling. Better health is an example: there is evidence that more highly educated individuals are better able to keep themselves healthy (see Haveman and Wolfe, 1984). Another example is the pure "consumption" benefit of knowing more about the arts and sciences. In addition to such benefits accruing to educated individuals themselves, other members of society also benefit. As mentioned above, we rely on each other's competence not only at work but also outside the sphere of market activity. And education enables some people to make contributions to the culture (e.g., basic research) whose value exceeds what the contributors are paid.

Computers in education and cultural transformation. Having reviewed basic economic notions of efficiency as applied to education, we can now consider the possible effects of computers. As Pogrow (1986) pointed out, we can judge the efficiency of computers for accomplishing current purposes in education -- e.g., raising scores on standardized tests -- or we can consider whether computers enable us to achieve completely

new purposes. For judging efficiency in light of current purposes, Levin (1983), Levin and Woo (1981), and Dalgaard et al. (1984) have specified the resource ingredients of CAI programs and costing procedures. A good example of cost-effectiveness study for a school district is Hawley, Fletcher, and Piele (1986). They find CAI is cost-effective relative to "traditional instruction" for teaching math skills to third and fifth graders. Outside of elementary and secondary schools, cost-effectiveness of computers for instruction has been studied most extensively by the military (Orlansky, 1985; Fletcher and Orlansky, 1986).

However, Linn (1986) has raised the possibility that such analysis -- or, indeed, any evaluation of computers in achieving traditional educational purposes -- may delay or deter progress toward realizing the most important new purposes of computers in education. Linn's first recommendation for evaluating computers -- and, more generally, "technology" -- in education is to identify and nurture new ideas to discover promising new paths, not to compare them with methods for achieving "traditional" purposes.

"It is clearly far more important to identify and build on creative paths for technological programs than it is to compare the innovations to traditional programs. Comparison of new programs with traditional ones fails to say why the new programs work. In contrast, evaluations of innovative programs should identify their strengths and weaknesses. Then existing programs can be improved and new programs can incorporate effective techniques.

"Comparing programs using technology to traditional curricula rarely yields useful information since new programs usually have goals attuned to technological change, efficiencies attributable to technology, revised roles for students, and new responsibilities for teachers...

"Evaluation rests, in part, on the value one assigned to these changes in the curriculum. Often these changes are highly valued and not actually open to revision. Essentially, administrators or teachers may have decided to teach students to use new technological tools, to increase the available problem representations, or to provide low income students with access to computers. The evaluation question concerns how to accomplish this change effectively, not whether a program that was just created does everything as well as the previous program while also achieving a myriad of new goals." (p. 13)

Furthermore, Linn argues that subjecting technology-based programs to conventional cost comparisons is a waste of effort.

"In general, there are so many differences between the two approaches that comparison is silly. Often, material taught in technologically-based programs simply can't be taught in another way, and therefore is cost effective by any criteria. For example, flight simulators teach skill in dealing with emergencies that cannot be created in another way. In another example, the music microworld developed by Balzano's FIPSE project teaches music composition skills that could only be practiced if students had full orchestras available to try out their fledgling ideas.

"In other cases, material taught in a technologically-based curriculum matches its application. It would be pointless to teach students to use paper and pencil spreadsheets and then expect them to apply this knowledge to a technologically presented spreadsheet, just as it would be pointless to teach technologically-based spreadsheets and expect students to use a paper and pencil approach. The skills required differ substantially.

"Other cost issues also require careful thought. If students will need computers for all of their activities, then the cost of purchasing hardware for students ought not be to considered for each project, but rather should be shared among them. Indeed, it seems likely that computers will become so widespread that students will purchase them as they purchased typewriters in the past.

"In summary, the cost effectiveness of technologically-based curricula defies straightforward analysis. Developing on inexpensive hardware is probably not cost-effective because as the hardware becomes even less expensive, the development effort will prove to apply only to obsolete machines and not take advantage of the potential of technology. To follow promising paths, developers need the flexibility of powerful machines. They may instead adapt developed programs to inexpensive machines.

"Comparing costs of traditional and technologically-based programs usually fails because the goals and objectives of the programs differ substantially." (pp. 21-22)

Linn and others point to these small ways in which computers already have extended our capabilities in education. But this is only the beginning, perhaps. Finding out is the point. Alan Schoenfeld (1986) concluded, "The main issue, then, is to see in what ways computer-based technology can or might effect fundamental changes" (p. 13). Using Pea's terms, Schoenfeld elaborates on "the notion of computers as transformers of cognitive functioning rather than . . . amplifiers of it." In the same vein, diSessa (1986) describes how the advent of technology to support artificial intelligence and cognitive science now gives us "a real shot at understanding and even simulating thinking and knowing in a fully scientific way" (p. 2). Such understanding "can turn the art of education into a principled

scientific and engineering enterprise with leverage in terms of educational achievement far beyond what we historically expect of an evolving but largely informal and inarticulate professional practice" (p. 30). These authors all emphasize that we are only at the beginning. We do not know what is possible, but if we are seriously interested in the future, we should invest some effort toward finding out.

At the same time, we would add, it would be unwise to ignore the consequences of current CAI programs for doing what we currently want schools to do: help students master written and oral communication (including math), acquaint them with important areas of human knowledge, and enhance their ability and motivation to solve problems. If schools were buying technology simply because it was cheaper than hiring more teachers, or perhaps because it made it easier for teachers to ignore some students some of the time, and if students were not learning as much of what we want them to learn now, we would not consider this a wise use of resources from either the long-term or short-term point of view. The exciting future potential of computer-based technology in education does not diminish our concern for careful use of resources to achieve current purposes. Cost-effectiveness comparisons of computer-based programs in education versus "traditional" practice, versus "innovations" not based on physical technology (e.g., students tutoring other students, or some form of "cooperative learning") are still appropriate.

There is also the possibility, described by Levin and Meister (1985), that computers will fail to fulfill their potential in education because they do not prove cost-effective for achieving "traditional" purposes. Motion pictures, radio, and television all were predicted to revolutionize education, but most elementary and secondary schools (both private and public) still make only limited use of these technologies. Most schools now have a computer for every one or two teachers -- just as they have movie projectors, radios, and televisions or video tape players in small numbers. But whether schools will ever acquire enough hardware and software so that all students can work on computers at the same time will depend in part on judgments of cost-effectiveness along the way.

Cost-effectiveness and decisions in education. Daily decisions governing resource allocation in schools are made by teachers, students, principals, district administrators and governing boards. Schools and districts also make annual budget decisions. States, foundations, universities, corporations, churches, and the federal government all have budgets for education. Decisions regarding these resources affect the development of computer-based technology in education. What kind of cost-effectiveness study is useful?

Operating decisions by teachers (e.g., which student to "call on" next), by students (e.g., which homework to do first), and by public school principals (which fires to fight today) are often about allocating time. There are so many of these decisions that the notion of doing cost-effectiveness study on each one is absurd. However, the existence of time constraints at the daily level has to be reckoned with in higher-level decisions about dollars. If part of the educational process becomes automated, what happens to the allocation of time under current circumstances? We will consider this further in the next section.

Where cost-effectiveness studies apply most readily is in decisions about allocating money. School sites and districts have some discretionary budgets. In the short run, the teaching staff and physical facilities change slowly, but the discretionary budget sometimes covers acquisition of hardware and instructional software. Such expenditures have brought us to where schools have approximately as many computers as teachers (Winkler et al., 1986). Should schools try to acquire as many computers as students? Some schools already have a computer for every student, and in many schools there are certain classrooms where every student is working on a separate computer or terminal. It is possible to find out how this has affected student learning, and to estimate whether equally good results could be achieved by less expensive means. It is also possible to compare effects of various instructional arrangements at the saturation density of one computer or terminal per student. In a subsequent section we discuss some of the major differences among instructional configurations.

State governments, which now pay the largest share of public school costs, have to balance claims for education against claims for highways, public welfare, health and hospitals, and natural resources. Given limited budgets for education, should money be spent on buying a lot more computers now, or waiting on computers and spending the money now to reduce class size, raise teachers' salaries, or sponsor special programs? If the political process were more accurately informed about cost-effectiveness, these decisions could be more rational. The State of Arkansas has shown how such information can begin to be assembled (McDermott, 1985).

Finally, the federal government currently spends money for vocational, compensatory, and higher education, among other things. Could these programs be more cost-effective if they encouraged more or better use of computer-based technology? Again, since schools and classrooms vary considerably in their use of computers now, a good deal could be learned from comparing current practices.

In sum, cost-effectiveness studies have a logical connection to decisions at local, state, and federal levels. It is important to remember, however, that cost-effectiveness can be measured only in relation to the current purposes of schooling. If development of intelligent tutoring systems and other products of research in computer science and artificial intelligence leads to new definitions of "what knowledge is" (diSessa, 1986, p. 3), then the purposes of education presumably will change along the way. Evaluating programs in terms of unknown objectives is beyond cost-effectiveness study. As explained above, basic research along these lines must be evaluated on other grounds.

Multiple Outcomes and Limited Time

One major complication in educational cost-effectiveness research is the existence of many outcomes. Given that students and teachers have limited time, any innovation that changes the use of time will generally have multiple effects. If additional time is spent on one particular subject, then less time must be spent on something else. Consequently,

students' performance may improve in the subject targeted by the innovation, but performance may decline in something else. One apparent case of such effects was the experiment in educational performance contracting conducted in 1970-71. Students taught by contractors achieved slightly bigger gains in the subjects on which payments to contractors depended, but students in control groups did better in non-contracted subjects (Gramlich and Koshel, 1975, pp. 36-37).

Even when incentives to emphasize certain outcomes are not as explicit as in the performance contracting experiment, they do exist. Innovations are usually designed to achieve certain objectives more than others. For example, it is possible to design a CAI package to teach a certain narrow set of skills, then test improvement in these skills to measure the effect of CAI. If the CAI program took a substantial amount of students' time but the evaluation did not also consider other outcomes, it would be quite misleading. Meta-analysis of such results would be misleading as well.

Glass discussed the issue of time in his analysis of the CAI study in Los Angeles conducted by the Educational Testing Service (Ragosta, Holland, and Jamison, 1982; part 6). The treatment in this study consisted of daily drill-and-practice sessions conducted in a lab outside the regular classroom. Sessions lasted 10 or 20 minutes. Within a given classroom, students were randomly assigned to receive CAI in math only, reading and language only, or a combination of math with reading and language. While students assigned to math CAI were at the lab doing math, other students from the same classroom would either be in the lab doing language or reading, or they would be back in the regular classroom doing whatever the teacher's plan called for: sometimes math, but more often other subjects, since math does not take up most of the regular day in elementary school. As a result, students assigned to math CAI would spend more time on math than other students. Glass estimated that students receiving 20-minute daily sessions in math CAI spent an extra 50 hours a year on math, over and above the approximately 90 hours a year of math instruction received by other students. That is a large increment of time.

Extra time in one subject implies less time for other subjects. If performance in these other subjects declined, the innovation would have to be considered less successful than if such a decline did not occur (or were ignored). Unfortunately, most evaluations of CAI have not reported any evidence, one way or the other, on performance in subjects not specifically taught by the CAI itself. The ETS-LA study is an exception, since all students were tested in a range of subjects. However, in the ETS-LA study, students who received CAI only in math served as the controls for students who received CAI only in language and reading. Therefore, the design of the ETS-LA study makes it difficult or impossible to distinguish between (1) students doing worse in subjects where they received no CAI (and therefore had less time allocated), as opposed to (2) students doing better in subjects on which they received CAI (and spent extra time).

Some innovations are more subject-specific than others. Tutoring, like CAI, usually focuses on one or two subjects. In contrast, the effect of reducing class size, extending the school day, or raising teachers' salaries could be spread over all subjects. Computer-managed instruction (CMI) systems also may have a diffuse impact on students' performance if their first effect is to save teachers' time. Test authoring programs, record keeping programs, word processors, and other computer-based tools may improve teachers' effectiveness in various ways, but up to now the effectiveness of such tools has been examined through relatively narrow measures of students' achievement gains. These tools, however, may be producing other kinds of effects, such as helping teachers maintain records with less effort, allowing them to create work sheets or tests faster, making forms neater and visually more appealing, so that teachers have more pride in their output. Effects such as these may not show up quickly in student achievement because they only indirectly affect students. The primary impact is on the teachers and should be observed at that level, using variables such as greater efficiency, greater pride in management tasks, less teacher burnout, or other appropriate indicators. Any effect on student learning will be relatively slow and diffuse.

Some innovations, including CAI and tutoring, require that students spend time in transition from one place or activity to another. Again, since this transition time detracts from instructional time in other subjects, it is important to evaluate effects on a broad range of outcomes, not just the particular subjects or skills taught by the innovation.

With regard to CAI, it should be mentioned that there is a definite possibility of saving students' time. "Virtually all studies show that students complete the material faster with CAI than with traditional instruction -- sometimes as much as forty percent faster" (Capper and Copple, 1986, p. 6). Kulik (1985) synthesized the results of 28 investigations of instructional time and found an average reduction of 32 percent. Such time savings mean that CAI can potentially teach students more without taking time away from other subjects.

Recognition of time constraints implies that an innovation should be evaluated on a broad range of outcomes if it changes students' use of time. If CAI or tutoring increases the amount of time students spend on a particular subject, and therefore reduces time on other subjects, then it is misleading to use performance in that subject alone as the criterion for comparing the CAI or tutoring program against "traditional" instruction, or against some other innovation that has a diffuse but potentially positive impact on performance in all subjects. Taken to an absurd extreme, it would be possible to have students spend all their time practicing a particular skill such as long division, then give a test on long division only, and find a large "effect size" compared to students who had not spent all their time on that narrow subject. To avoid such misleading conclusions, evaluations have to consider the whole range of subjects students are supposed to be learning.

Outcome measures in existing evaluations of CAI. Individual studies of CAI generally have not examined a broad spectrum of outcomes simultaneously. However, various outcomes have been considered by different individual studies, so that the research as a whole does include an array of outcomes. The most common technique for assessing CAI has been through the use of subject-specific achievement tests, either nationally

normed or locally developed. Based on a recent series of meta-analyses (Bangert-Drowns, Kulik and Kulik, 1985; Kulik, Kulik and Bangert-Drowns, 1986; Kulik and Kulik, 1985), Kulik has concluded that "students have generally learned more in classes when they have received help from computers" (Kulik, 1985, p.3). Niemiec and Walberg (1987) synthesized research syntheses by a number of other reviewers, with the following results:

<u>REVIEWERS</u>	<u>MEAN EFFECT SIZE</u>
Aiello and Wolfle	.42
Bangert-Drowns and others	.42
Burns and Bozeman	.39
Hartley	.42
Kulik and others (adult)	.42
Kulik and others (college)	.26
Kulik and others (elementary)	.47
Niemiec and others	.45
Samson and others	.32
Schmidt and others (special education)	.56
Shwalb and others	.45

"Higher-order" thinking skills are a more elusive, but perhaps also more important, outcome. Examination of this outcome is a very difficult evaluation problem because there is no universally accepted description or definition of what higher-order thinking skills are or how to assess students' competence in this area. Research on use of computers to develop higher-order thinking has not yet produced any major breakthroughs, but the possibilities are exciting. Pogrow (1986a, p. 22) states that "technology at present is incapable of providing the sophisticated forms of interaction needed to develop thinking. It can only tell a student whether an answer is right or wrong -- not why an answer is wrong

and then help remediate that specific problem." But Pogrow himself is currently developing and distributing a curriculum with the primary goal of using the computer to create learning environments that will "develop the general information processing skills that lead to more sophisticated information networks in long term memory," networks which "are essential to support generalization and retrieval abilities critical to problem solving in all content areas." Other researchers have shifted their emphasis from general mechanisms and universal processes for problem solving to cognition in specific domains, e.g., mathematics. Their understanding is that problem solving strategies and heuristics are domain-specific, and that within this context "The main issue is to see in what ways computer-based technology can or might effect fundamental changes..." (Schoenfeld, 1986). At this reserach frontier, basic theories of knowledge are evolving hand-in-hand with educational applications (diSessa, 1986).

A third set of outcomes have to do with the affective impact of educational technology. This area, however, has received relatively little attention. In the series of meta-analyses carried out by Kulik et al., of 199 studies, only 17 examined students' attitude toward computers and toward instruction, while 29 measured students' attitude toward the subject matter being taught. Generally, Kulik found that students' attitudes toward computers and toward instruction improved with the use of CAI. Attitude toward subject matter, however, varied across the 29 studies resulting in an average effect size near zero.

The importance of affective outcomes was highlighted in three case studies of microcomputer use carried out by Sheingold, Kane, and Endrewert (1983). They found that "teachers report primarily the social outcomes related to interaction, status, and self-esteem." Pogrow's (1986b) discussion of his higher-order thinking skills ("HOTS") curriculum mentions several affective advantages of a computer-based learning environment. These include: enhancing the students' motivation and willingness to think, that computers signal a different learning environment which students do not associate with

previous classroom experiences, and that the computer oriented environment of "learning from visual images" provides "an island of familiarity in a very strange world." Similarly, Schoenfeld (1986) describes how computer-based instruction may influence students' attitudes and beliefs toward a subject (e.g., math), their sense of what a subject is about, and the importance and credence they give to conclusions or propositions in that subject.

Introduction of computers into classrooms can influence a wide array of student outcomes because computers lead to various changes in teachers' use of class time and other resources. Clark (1985, p. 140) observed that "Designing computer-based instruction programs is taxing and requires more systematic thinking about information organization, examples, and requirements for active responding," thus requiring teachers to do more planning. The case studies by Sheingold et al. (1981) found that introduction of computers resulted in curriculum being rewritten and in classroom organization and relationships changing -- even though this had not been the specific intent. Shavelson et al. (1984) emphasized the importance of such changes in defining instructional use of computers as:

"the appropriate integration of computer-based learning activities with teachers' instructional goals and with ongoing curriculum, which changes and improves on the basis of feedback that indicates whether desired outcomes are achieved." (p. 25)

Shavelson's group described how computers influence time allocation, what strategies are used for assigning and grouping students for various CAI activities, and what rules are established concerning computer use. In another study, a group of junior high teachers found that using CAI to teach basic skills allowed them to focus their own efforts on teaching higher-level thinking skills (Carney, 1986). The State of California, recognizing the stimulation educational technology can provide to educational practice, recently published a series of resource guides, Technology in the Curriculum (California Department of Education, 1986), that encourage teachers to use computers and related technology for improving the curriculum generally.

Time constraints on teachers also imply that the benefits of computers in schools do not accrue only to students who use them. While some students are busy on computers, other students can receive more attention from teachers. Like books, "dittos," science corners, and other "activity centers," computers provide an additional option for occupying some students while teachers work with others.

For these reasons, previous studies of computers in schools may not provide accurate measures of their impact. Most of these studies have ignored the possible effects of reallocating students' time on their performance in subjects other than those specifically taught by CAI. Some studies have considered higher-order thinking and affective outcomes, but most have not. And most previous studies have not accounted for possible indirect benefits to students when computers are introduced in schools. To get a more accurate assessment of computers' effects in schools, each individual study should measure a range of outcomes for students.

Cost-effectiveness and multiple outcomes. When multiple outcomes are considered, analyzing cost-effectiveness becomes more complicated. For instance, suppose two programs are to be compared: CAI and a cross-age tutoring program. Suppose that there is a control group for each program, and a broad spectrum of outcomes are measured for students in each program and each control group. Suppose further that CAI is found to produce bigger gains than the tutoring program in math and in students' motivation to do schoolwork, but the tutoring program produces bigger gains than CAI in reading and in higher-order thinking. To make the comparison as simple as possible, suppose the two programs have equal cost. Which is more cost-effective?

One approach to this problem is to combine different outcomes into a single score. If the combination procedure is linear, this means assigning fixed weights to the different outcomes, giving more weight to outcomes that are considered more important. The resulting score then can be treated as a single outcome, and cost-effectiveness comparisons can proceed as usual. This is what Levin (1983) calls "cost-utility analysis."

Assigning weights to different outcomes requires, first, a procedure to select judges and, second, a procedure to elicit their judgments about the relative values of different outcomes. The first procedure is political. If the weights attached to different outcomes are to represent preferences of some constituency, representatives of that constituency have to be included. The constituencies interested in public schools include parents, teachers, students, administrators, taxpayers, and employers. A legitimate set of weights on different outcomes would have to represent these different groups. The judges who assign these weights would have to be selected accordingly.

Once a sample of judges is chosen, there are various possible procedures for determining weights. A simple procedure would have each judge assign a number -- from one to ten, one to a hundred, or whatever -- to each outcome, where a larger number signifies greater importance. These numbers can then be simply averaged across judges to derive weights.

An example of a more complicated procedure for deriving weights is "conjoint analysis" (see Rao, 1977). This has been used in marketing research to discover what differences between competing brands are important to consumers. A sample of respondents is asked to make hypothetical choices -- in this case, the choice would be among a set of hypothetical programs. Each program would be described in terms of outcomes: high on some, low on others. Respondents' ratings of programs can be analyzed to provide numerical weights signifying each outcome's importance. Results would probably be more accurate, in the sense of predicting how people choose schools, than would results from the simpler weighting procedure. However, conjoint analysis is usually done with only three or four dimensions to be considered. In education, it would be difficult to reduce the number of dimensions to less than a dozen.

Whether simple or sophisticated, procedures for eliciting weights are inevitably artificial and somewhat arbitrary. Arbitrariness is also a problem in selecting the judges. In practical terms, the resulting weights are likely to be unstable. If new weights are

generated in subsequent years, they are likely to cause changes in the cost-effectiveness rankings of different programs -- changes that are not due to changes in actual performance. Arbitrary and unstable weights will not guide a steady effort toward educational improvement, but, instead, will promote fickle and superficial changes.

A final drawback of these weighting schemes is that they should be non-linear, but existing procedures produce only fixed weights for making linear combinations. Non-linear combinations are theoretically preferable because the value of a one-point gain in a particular outcome presumably is greater if students are performing at a low level than if they are already at a high level ("diminishing marginal utility"). Also, the value of a one-point gain in one outcome may be greater if students are performing at high levels on other outcomes than if the other outcomes are low ("diminishing marginal rate of substitution"). However, it is difficult to elicit statements of preferences that satisfy these theoretical criteria.

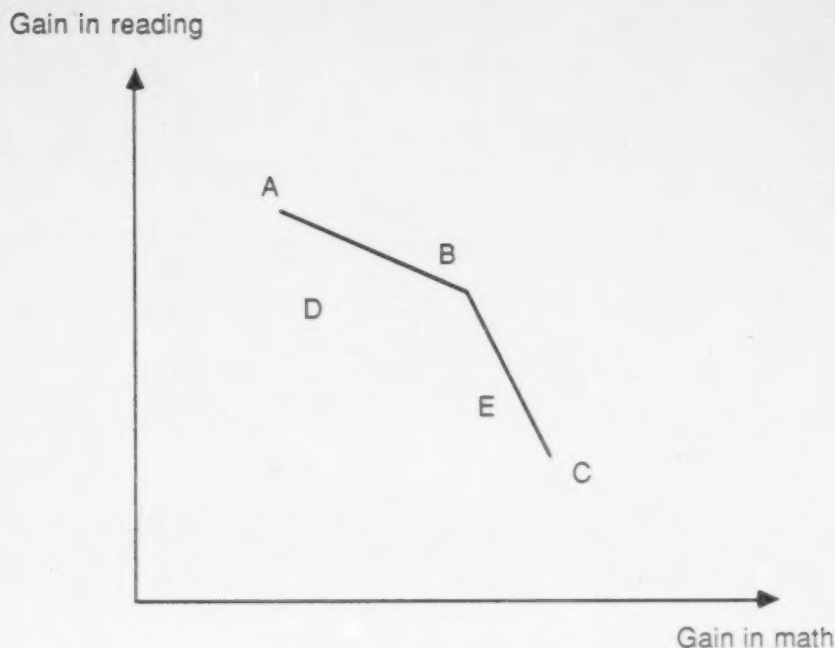
There is an alternative to analyzing cost-effectiveness which takes multiple outcomes into account but does not impose arbitrary, fixed weights on different outcomes. Charnes, Cooper, and Rhodes (1978) have developed a linear-programming approach to this problem, which they call "data envelopment analysis" (DEA). As they have described it, the purpose is to determine which "decision-making units," among a set of similar decision-making units, can be considered efficient. Schools are an example of decision-making units. Bessent and Bessent (1980) have, in fact, demonstrated the empirical application of DEA to schools in Houston, Texas.

DEA begins with the idea of an efficiency index, which can be computed for each decision-making unit. The index is the ratio of a weighted sum of outputs to a weighted sum of inputs. A different set of weights is computed for each decision-making unit, to maximize its efficiency index (constrained not to exceed unity). If the maximum efficiency index computed for a given decision-making unit, using the set of weights that make it look as efficient as possible, is less than the efficiency index computed for another decision-

making unit using the same set of weights, then the first decision-making unit is deemed relatively inefficient.

If the purpose is to compare programs such as CAI or tutoring, instead of comparing decision-making units such as schools, then DEA could be modified as follows. Suppose we have data on dozens or hundreds of programs. Each program is some kind of intervention or alteration in a school or set of schools. From each program the data include a set of outcomes for students in the program compared to a control group. Suppose we also know the cost of each program. Then, using an algorithm like DEA, we could compute the weights on different outcomes that would yield the largest possible efficiency index (weighted effect size per dollar) for each program. A program would be considered inefficient if its efficiency index, using its "own" weights on different outcomes, is smaller than the efficiency index of some other program, computed with the same weights.

The following figure illustrates what could be learned from this procedure. A, B, C, D, and E are five programs. For simplicity, we consider only two outcomes, and assume here that all five programs cost the same amount of money per student. Among the five programs, A, B, and C would have maximum efficiency ratings, but D and E would not. The lines connecting efficient programs represent what economists call the production possibilities frontier. Programs A, B, and C are on the frontier, but D and E are not. We have not worked out the actual linear programming algorithm for this analysis, but it would seem to be a fairly straightforward extension of DEA.



Procedures like DEA have certain drawbacks. One problem is that results might change drastically if only one or a few observations were not included in the analysis. For instance, in the diagram above, program E would appear efficient if program C were not included. In addition, errors in measurement can produce large distortions in DEA results. Sexton, Silkman, and Hogan (1986) discuss these and other drawbacks of DEA, and suggest possible modifications of the technique. Standard economic procedures for estimating production functions may be more robust against errors in data or model specification, but these standard techniques do not handle more than one output at a time (see Forsund, Lovell, and Schmidt, 1980; also Levin, 1974). It is possible to handle multiple outputs using a cost function instead of a production function (e.g., see Cavin and Stafford, 1985). If the different educational programs being analyzed pay different prices for a given resource such as teachers' time, then estimating a cost function would be a feasible alternative to DEA as a technique for measuring efficiency.

In sum, it is possible to define and measure cost-effectiveness when multiple outcomes are considered. This would enable us to examine educational innovations or interventions at different times and places, and to decide which ones have been relatively efficient. The result might be that some CAI programs appear efficient and others do not, some tutoring programs are efficient while others are not, and similarly for other types of programs. Generalizations about the cost-effectiveness of one type of program (e.g., CAI) relative to another (e.g., tutoring), would be built upon analysis of many specific instances.

Synthesizing Research on Multi-Dimensional Treatments: The Case of CAI

In addition to considering multiple outcomes, educational evaluations also have to deal with treatments that are complex and multi-dimensional. This is certainly true in evaluating CAI. As Clark (1983) has argued, estimating effects of the computer per se requires separating the medium from the method of instruction: e.g., comparing the effects of drill-and-practice with computers versus drill-and-practice with human teachers. However, this is difficult in practice because there are so many different dimensions for classifying CAI programs. Here we describe some of these dimensions.

One of the commonly discussed dimensions of CAI interventions is the type of software used. Many different ways of labelling educational software exist, but the usual categories are drill-and-practice, tutorials, simulations, languages, tools (including word processors, databases, and laboratory instrumentation), and computer-managed instruction packages (CMI) such as test authoring programs and record keeping programs. It is generally well recognized that these are all very different kinds of computer-based instruction. An example of this is drill-and-practice versus CMI. These two types of software are, respectively, analogous to practice worksheets and a teacher's grade book in a traditional classroom. The results that could be expected from two such interventions would be very different.

Another major dimension of CAI interventions is the degree of separation or integration with regular curriculum and instruction. Shavelson et al. (1984) conducted a study of 60 "successful" microcomputer-using teachers and their patterns of microcomputer-based instruction. The study found four major patterns of use, distinguished by the types of software used and the degree of integration with regular curriculum and instruction. This study emphasized the importance of these two dimensions of computer-based instruction treatments.

A third major dimension is individual versus group use of CAI. As mentioned above, the teachers participating in Sheingold, Kane, and Endreweit's case studies "reported primarily social outcomes related to interaction, status, and self-esteem," but such results cannot be found in treatments where students work individually. A feature of the "HOTS" curriculum being developed by Pogrow is that several students start out working together on a unit and as time passes activities become individualized. This structure "reduces anxiety at having to learn something new by oneself, and provides students with an opportunity to learn how to cooperate and communicate around the use of ideas" (Pogrow, 1986b, p. 61). Additionally, a study by Johnson, Johnson, and Stanne (1986) comparing computer-assisted cooperative, competitive, and individualistic learning found that:

"Computer-assisted cooperative learning promoted greater quantity and quality of daily achievement, more successful problem solving, more task related student-student interaction, and increased the perceived status of female students." (p. 382)

The National Survey of School Uses of Microcomputers (Becker, 1983/84) found significant variation between schools, grade levels, and subjects in the way students are organized for computer use and in the way time was used by students and teachers while others were working on computers. These findings indicate that the group/individual dimension is relevant to the effectiveness of CAI in achieving a broad range of outcomes.

Another dimension of computer-based instruction interventions which should be considered is subject area. This is a commonly used dimension in literature reviews and meta-analyses. Niemiec, Samson, Weinstein, and Walberg (in press) conducted a meta-analysis in which they identified course content as: unspecified, vocabulary, comprehension, language arts, spelling, math problem solving, math reasoning, social studies, total reading, total math, science, and other. This taxonomy identifies research studies not only by subject area, but also by skill areas, e.g., math problem solving, vocabulary. Another taxonomy for this dimension is provided by the meta-analysis of Bangert-Drowns, Kulik, and Kulik (1985). This study identifies subject areas of mathematics, science, social studies, language/reading, other, and combined (involving course content from several subject areas). These two analyses found that the effectiveness of CAI varies from one subject to another.

The ability level of students for which CAI programs are designed and implemented is also an important dimension. It has been documented that computer-based instruction has different levels of effectiveness for students at different ability levels (Bangert-Drowns, Kulik, and Kulik, 1985; Fisher, 1984). This dimension may also interact with the dimension of software type. It has been observed that lower ability students tend to receive more drill-and-practice than other types of software, while higher ability students receive more simulations and enrichment activities (Shavelson et al., 1984; Niemiec et al., in press). The result of the confluence of these two variables is not known; their interaction may increase effectiveness, inhibit effectiveness, or do nothing at all.

A dimension of CAI interventions that has only begun to be investigated is computer access. This is not referring to "equity of access" to computers for various groups of students, but to classroom organization for access to the computer and courseware. Shavelson et al. (1984) reported that more than a third of the sixty teachers involved in their study had rules governing operation of microcomputers and about talking while at the computer. Only approximately one-fifth of the teachers had rules concerning time at the

computer. They also found considerable variation among teachers' estimates of the number of minutes per week students received instruction on the computer, but the average was 51 minutes per week.

In comparison, the national survey of microcomputer use carried out by Becker (1983/84) found that "the typical micro user in a junior high school gets 30 minutes of access time per week... half as much as the typical high school student user gets and not much more than the 24 minutes per week typical of elementary school users." (It should be stressed that the "typical micro users" in this study are a minority of students in any one school.) This survey also found that free time use of microcomputers (before school, after school, and during lunch) averaged 4.6 hours per week in senior highs, 2.5 hours per week in junior highs, and 2.2 hours per week in elementary schools, for each school-owned microcomputer. A factor that forms part of computer access is the location of computers. Computer-based instruction is now carried out in a number of different situations. Microcomputers are commonly found in computer labs, media centers, individual classrooms, or rotating in and out of several classrooms. It is easy to see how location of computers can influence students' access to computers and how computer-based instruction is integrated into the regular curriculum and instruction.

These initial investigations show that there is a great deal of variation in how restricted students' use of computers is or if students have free use of the computer, but we do not yet have any systematic knowledge about which configurations are most effective.

One final dimension we would emphasize, which has not been treated in existing literature reviews and meta-analysis, is whether CAI focuses on basic skills, higher cognitive skills, computer literacy, or attitudes. This dimension has been hinted at by such factors as subject area and software type, but not directly addressed. Yet awareness of objectives seems critical for the proper assessment of CAI. The Autonomous Classroom Computer Environments for Learning (ACCEL) project demonstrates this. ACCEL "is designed to isolate and test instructional provisions that foster autonomous learning" (Linn,

Rohwer, and Thomas, 1986, p. 2), and this is to be done by fostering ideal programming knowledge. ACCEL might be described as a computer literacy program whose objectives are higher cognitive skills. In general, any program designed for a specific content area may be directed at basic skills or at higher cognitive skills. Full understanding and proper assessment of the effectiveness of computer-based instruction interventions would be enhanced by consideration of this dimension.

We have described several major dimensions of CAI: type of software, degree of integration with regular curriculum and instruction, individual/group use, subject area, student ability level, organization of computer access, and focus of objectives. All of these are part of, or concomitant with, every actual CAI intervention. Clark's desideratum "that variables other than the media that are competing be held constant" is not attainable in the field. Even if it is possible to create research situations in which all variables are held constant, this would not produce educationally meaningful results.

Meta-analytic approaches. Several meta-analyses of research on CAI have been done (for recent examples, see Bangert-Drowns et al., 1985; Niemiec et al., in press; Roblyer, 1985). The unit of observation in a meta-analysis is a sample of students for which some outcome (usually only one) can be compared to a control group. The mean difference in outcome between the sample and its control group, divided by the standard deviation within the control group, is usually taken as the measure of "effect size" (Glass et al., 1981). A single research study may report results for more than one sample, but most meta-analyses have given less weight to multiple results reported by one study than if they had been reported by separate investigations.

Meta-analyses of CAI have examined the relationship between effect size and certain characteristics of samples or studies. For instance, Bangert-Drowns et al. (1985) analyzed 42 studies of computer-based education in secondary schools. They coded each study on 16 different features, including duration of treatment, course content, and whether the treatment and control groups had the same or different instructors. Similarly, Niemiec et

al. (in press) analyzed 45 features of 48 studies on computer-based education in elementary schools. They grouped study characteristics into five categories having to do with the study document (e.g., year of publication), students (e.g., grade level, ability), predictor variables (e.g., duration of treatment, content area), dependent variables (e.g., type of test used to measure outcome), methodology (e.g., sample selection and method of data analysis), and threats to valid inference (e.g., reliability of measures).

Statistical procedures in these meta-analyses have consisted of one-way analysis of variance: comparing effect size by study characteristic, one characteristic at a time. This kind of analysis has found, for example, that studies of drill-and-practice at the elementary level have reported bigger effect sizes than studies of problem solving or tutorial software (Niemiec et al., in press). Such findings do not control for other characteristics of studies. For instance, would we find the same result about drill-and-practice if we took into account year of publication and ability level of students at the same time?

Answering this kind of question requires a multivariate procedure, such as multifactor analysis of variance, or multiple regression. Ordinary least-squares regression has been used in meta-analysis on the effects of class size, television, psychotherapy, and other topics, but not in meta-analysis of CAI. With the number of CAI studies constantly growing, it should be possible to use multivariate methods.

Hedges and Olkin (1985, chapter 8) have developed a theoretically attractive method for meta-analysis using weighted least-squares regression. They recommend using as the dependent variable an unbiased estimator of effect size in which the denominator is the standard deviation of the outcome for the treatment and control groups combined, rather than the control group alone. The error term in the regression is the difference between the estimated effect size in a particular study and its true value, which is assumed to be a function of observed study characteristics. Their theoretical derivation of the variance of this error provides the basis for the weighted least-squares estimation procedure. Although the statistical theory gives large-sample results, Hedges and Olkin report a Monte Carlo

study indicating the procedure is quite accurate for synthesizing results of studies even when the number of studies to be synthesized is small.

New Possibilities and Exemplary Programs

Multivariate meta-analysis and DEA are logical next steps for analyzing relative cost-effectiveness of existing or recent programs. As part of its assessment effort, OTA could assemble the necessary data base, perhaps drawing on recent meta-analyses of computer-based instruction, peer tutoring, and other kinds of programs.

An obvious limitation of such analysis is that the data are all from the past. For instance, even recently published meta-analyses of CAI are based mainly on programs that used minicomputers with terminals, not microcomputers. It will be a few years before we have enough documented experience with micros to meta-analyze. Yet, stand-alone microcomputers (with greater capacity than today's micros) could well be the wave of the future in schools. Since computer technologies change very rapidly, accurate assessment of their potential in education requires the most current possible evidence. Therefore, in addition to careful analysis of evidence from the recent past, OTA should consider collecting first-hand information on projects that are currently breaking new ground.

Study of path-breaking programs can be useful not only for predicting the near future, but also for understanding how innovations occur in schools. At the beginning of this paper we mentioned the lack of incentives for schools to be efficient. Absence of incentives tends to impede systematic innovation. The process of innovation, and of diffusing successful innovations from their original sites, is also made difficult in education by the idiosyncratic nature of teaching and learning. As Murnane and Nelson (1984) have argued, research has not yet accounted for most of the difference between effective teachers, schools, or programs and less effective ones. The importance of personalities and local contexts means that we cannot construct readily transferable "blueprints" for effective education. Yet, in spite of these difficulties, innovation does occur. In fact, the

increased availability of microcomputers has stimulated a great deal of innovation by classroom teachers themselves. OTA could make a valuable contribution by shedding light on the process of innovation, if possible including a look at spontaneous grass-roots experimentation as well as large-scale, formal R & D projects.

Some exploratory research of this kind has been conducted by Shavelson et al. (1984). They studied 60 teachers who were recommended as successful in developing ways to use microcomputers for math and science instruction. Shavelson et al. identified four distinct patterns of microcomputer use among these 60 teachers. They labelled these patterns orchestration, enrichment, adjunct instruction, and drill-and-practice, defined as follows:

"Orchestration represented the widest variety of instructional activities closely linked to regular curricular activities... Enrichment capitalized on available courseware to familiarize students with microcomputers and, as a consequence, integrated microcomputer activities least with subject matter and other classroom activities... Adjunct instruction, emphasizing cognitive goals in a catch-as-catch-can manner that was not as closely integrated with the ongoing curriculum... Drill and Practice stressed this instructional mode and basic-skill objectives, and tied microcomputer-based instruction closely to the curriculum and classroom activities." (p. vii-viii)

This typology of instructional use of computers, even though derived only from math and science instruction, describes the various practices of teachers who are regarded as successful computer-users. We can also use this typology to describe other current efforts aimed at advancing the state of the art.

One such effort is the California Department of Education's Technology in the Curriculum (1986). This is a series of four "resource guides," one for each of four subjects: history/social science, language arts, mathematics, and science. These guides encourage the "orchestration" pattern of microcomputer use. They identify favorably evaluated computer software and instructional video, matching them with teaching objectives as established by the official state of California curriculum framework. Also, the guides provide information to teachers for answering two questions: "Why should I use

computer and video in my classroom?" and "How can I integrate these materials into the curriculum?" Anecdotes, examples, and model units are given.

Shavelson's term "orchestration" applies to the Technology in the Curriculum guides because they describe how to integrate a wide range of technologies into the existing curriculum. Many types of software -- simulations, drill-and-practice, tools -- are recommended. Different grouping strategies are described, including individual, three or four person teams, or teacher discretion. In each of the four subject areas, each model lesson contains follow-up activities that integrate the technology-based instruction with other learning activities. The Technology in the Curriculum guides are not specifically geared to students of different abilities; they are most suitable for the typical student, rather than the exceptional student. Lessons are designed to be adaptable to whatever situation exists in the schools, instead of requiring particular facilities. Both basic skills and higher-order thinking skills are addressed, as these are expressed in the California curriculum framework. These materials also express a goal of improving students' attitudes toward the subject matter, but this affective goal is not stressed as much as cognitive outcomes. By orchestrating various technological means for a variety of educational ends, these guides demonstrate how technology can be worked into the "score" of the classroom.

A very different approach to computer-based instruction is represented by the Computer Curriculum Corporation (CCC). CCC provides programs for math, reading, and language arts. There are several contrasts between Technology in the Curriculum and CCC. The CCC program uses only drill-and-practice software and it is of a highly individualized nature. It covers three subject areas, but each of these is treated separately. Each area can be well integrated with the elementary school curriculum. The three areas cover most topics in math curricula and common topics in reading and language arts. Designed as a supplemental drill-and-practice program, CCC software can be used with students of different abilities, but typically it is lower-ability students who benefit most from such material. Computer access in this program is controlled. The software runs on

a minicomputer with terminals set up in a computer lab. Students are rotated through the lab for ten or twenty minute sessions. The entire program is structured to help students master basic skills in any of the three subject areas. The CCC program exemplifies the drill-and-practice pattern of computer use.

An effort that is currently breaking new ground is the HOTS (Higher-Order Thinking Skills) program developed by Pogrow. This program is an extension of what Shavelson et al. call adjunct instruction. The HOTS program is not directed at a particular subject area. "The goal of using software in computer-involved environments is rarely to provide direct practice in a specific content skill, but to help indirectly set a state where the student can perform the necessary feats of thinking" (Pogrow, 1986b, p. 25). As a side effect, this program also expects to affect students' basic skills and social confidence in a positive manner. This program has selected several types of off-the-shelf software and structured them into a learning environment. This environment is a computer lab situation in which each student is to spend a minimum of 35 minutes at least four days a week. It is directed specifically at disadvantaged, lower-ability students, but is being generalized to other populations. This program represents an improvement over the adjunct instruction found by Shavelson et al. in that it is not "catch-as-catch-can," but a well-planned approach emphasizing cognitive goals.

These three programs all represent efforts to increase the effectiveness of computer-based instruction above what is typically found in schools now. Evaluation results of the two programmatic interventions, CCC and HOTS, has found them to be effective, compared to traditional instruction (Rogosta et al., 1983; Pogrow, 1986b). All three efforts are attempts to promote the positive evolution of current practice.

Other current efforts are attempting to lay the groundwork for more fundamental change. One such effort is the development of a complete course of study for a subject area, e.g., math or history, that is totally computer-based. Such an effort has been suggested by Levin and Meister (1985); an example is currently being developed by

Suppes and others. Another possibility lies in the further development of educational environments. In this area, Logo is well known. Logo is software that is not "complete and targeted to a particular niche in a school curriculum, but instead, it is open to curriculum developers, teachers, and students to find their own ways of using it according to their own goals and local needs" (diSessa, 1986, p. 24). Presently, diSessa is developing "a new kind medium" called Boxer. It is "like written language but with much extended interactive capabilities. Looking a little farther into the future, Boxer might well be the right 'soft' medium for distributing educational materials. Thus, a student would get an optical disk full of text, data, computational tools, simulations, ..., all combined in an interactive medium that he or his teacher can personalize, change, and add to" (diSessa, 1986, p. 28). Efforts like these are attempting to extend the frontiers of educational practice in revolutionary new ways.

The problem of software. One problem that may inhibit progress unless it is solved has to do with the economics of software production and distribution. Levin and Meister (1985) found software producers view illegal copying as a "horrendous problem." Pogrow (1986a) is also concerned about "multiple loading." The problem is simply that software can be read into one machine, then transferred electronically to another machine. Another user can thus be added at no cost. This violates normal licensing and copyright agreements.

The problem is analogous to what economists call "natural monopoly." For instance, a highway bridge crossing a river creates a natural monopoly. The distinguishing feature of natural monopoly is that a large initial cost cannot be recovered if customers pay only the marginal cost of the service to them. But it is inefficient to charge users more than marginal cost, because that deters some potential buyers from using services for which they are willing to pay the true marginal cost. The usual solution to natural monopoly is regulation, subsidization by government, or both.

In the case of computer software, illegal copying adds to the problem of publishers recovering their initial costs. If publishers raise prices, they encourage more illegal copying. This is a vicious cycle that could impede the full development of educational software. The problem seems to be worse for software than for other educational materials such as textbooks, because software can be copied more cheaply than books. Also, the initial cost per unit may be higher for software than for books. OTA could look into this. If software publishing really does have the characteristics of a natural monopoly, then some kind of subsidization would be appropriate, as Levin and Meister (1985) suggest.

Richard Nelson and his associates (1982; also Nelson and Langlois, 1983) have considered the general question of how the U.S. government can support industrial research and development. They found that federal support for R & D has been successful when the federal government itself has been a purchaser of the new products or technologies being developed, or when the R & D being supported has been at the same time basic enough not to threaten proprietary interests and applied enough to have practical applications, or when the R & D has been focused on applied concerns in an atomistic industry (namely, agriculture). This provides a context for considering possible approaches to subsidizing educational software development.

Implications for OTA Study

In order to learn as much as possible from existing evidence about the relative cost-effectiveness of computer-based technology in education, OTA should consider conducting or sponsoring a more comprehensive analysis than has yet been done. This comprehensive review would synthesize results from existing studies on computers in education, as well as tutoring programs, lower class size, longer school days, higher salaries for teachers, and other attempts to improve results in schools. Each study included in this review should contain information on a broad range of student outcomes, including achievement in several academic subjects, higher-order cognitive skills, and attitudes about education. Each study

also should include information on the per student cost of treatment --or, better still, on the amounts of resources used and the price of each resource (e.g., teachers' time). Feasibility of conducting this review depends on how many existing studies meet these requirements. If the number of studies exceeds the number of outcome dimensions, a method like Data Envelopment Analysis can identify programs or treatments that are relatively cost-effective. In addition, multivariate meta-analysis can measure associations between characteristics of programs or treatments and their relative efficiency.

In order to learn about the unfolding potential of computer-based technology in education, OTA should consider sponsoring or conducting detailed case studies of promising practices. Cases may range from practices which have already been commercially packaged and implemented in numerous schools to ideas still in the experimental stages. Promising practices in corporate or military training, and in higher education, also may be included. The point is to find out where the existing frontier is being pushed back, and also what new dimensions of educational attainment are becoming evident. In the course of these case studies, it would be useful to start developing a set of standard definitions and information protocols. In particular, information should be collected on a broad range of educational outcomes: specific skills and knowledge, higher-order cognition, and attitudes. Also, cost should be accurately measured. By setting standards for data collection, OTA can help ensure that our understanding of educational effects will not be hampered by lack of critical information as has been true in the past.

OTA might also go a step further and develop suggested reporting standards for all federally supported educational programs. Arkansas is an example of one state moving in this direction on its own (McDermott, 1985). The purpose would be to demonstrate how Congress can obtain greater benefit from educational research, by ensuring that each federally supported program collect comparable data on a broad range of educational outcomes, and on costs. Studies that measure only a single outcome, and that fail to report

costs, do not contribute toward building a data base that is adequate to support judgements about educational cost-effectiveness.

Finally, realizing the potential of computer-based technology in education will depend on continued development of new and better software. However, there are signs of market failure in the software industry. Specifically, average cost may typically exceed marginal cost for software, unlike textbooks. If so, educational software publishing may not be commercially viable. OTA should consider looking into this.

Bibliography

- Alkin, Marvin C. (1969). Evaluating the Cost-Effectiveness of Instructional Programs. Center for the Study of Evaluation of Instruction, University of California, 1969.
- Arrow, K. J. (1973). "Higher Education as a Filter." Journal of Public Economics, 2, pp. 193-216 (July).
- Ash, Maureen S. (1985). Meta-Analysis and Research Integration of Regression Studies. Berkeley, CA: University of California, doctoral dissertation in Education
- Bangert-Drowns, Robert; Kulik, James H.; and Kulik, Chen-Lin C. (1985). "Effectiveness of Computer-Based Education in Secondary Schools." Journal of Computer Based Instruction, 12:3, pp. 59-68 (Summer).
- Becker, Henry Jay (1983/84). "School Uses of Microcomputers." Journal of Computers in Mathematics and Science Teaching, 3-4.
- Becker, Gary S. (1975). Human Capital (2nd ed.). Chicago, IL: University of Chicago Press.
- Benson, Charles S. (1978). The Economics of Public Education (3rd ed.). Boston, MA: Houghton Mifflin.
- Bessent, Authella M. and Bessent, E. Wailand (1980). "Determining the Comparative Efficiency of Schools Through Data Envelopment Analysis." Educational Administration Quarterly, 16:2, pp. 57-75 (Spring).
- Boardman, Anthony E.; Davis, Otto A.; and Sanday, Peggy R. (1977). "A Simultaneous Equations Models of the Educational Process." Journal of Public Economics, 7:1, pp. 23-49 (February).
- California Department of Education (1986). Technology in the Curriculum. Sacramento, CA: California State Department of Education.
- Capper, Joanne and Copple, Carol (1986). "Computers in Education." The Research Into Practice Digest, 1:3 (Spring).
- Carnegie Forum on Education and the Economy (1986). A Nation Prepared. Teachers for the 21st Century. Report of the Task Force on Teaching as a Profession. New York: Carnegie Corporation.
- Carney, Cathy Chmielowski (1986). "Teacher + Computer = More Learning." The Computing Teacher, 13:6, pp. 12-15 (March).
- Cavin, Edward S. and Stafford, Frank P. (1985). "Efficient Provision of Employment Service Outputs: A Production Frontier Analysis." Journal of Human Resources 20:4, pp. 484-503.
- Charnes, A.; Cooper, W. W.; and Rhodes, E. (1978). "Measuring the Efficiency of Decision Making Units." European Journal of Operational Research, 2:6, pp. 429-444 (November).

- Clark, R. (1983). "Reconsidering Research on Learning from Media." Journal of Educational Computing Research, 53:4.
- Clark, R. (1985). "Confounding in Educational Computing Research." Journal of Educational Computing Research, 1:2, pp. 137-148.
- Cohen, D. K. and Murnane, R. J. (1985). "The Merits of Merit Pay." The Public Interest, 80, pp. 3-30.
- Cohn, Elchanan (1979). The Economics of Education (revised ed.). Cambridge, MA: Ballinger.
- Coleman, James S. et al. (1966). Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office.
- Committee for Economic Development (1985). Investing in Our Children. New York: author.
- Dalgaard, Bruce R.; Lewis, Darrell R.; and Boyer, Carol M. (1984). "Cost and Effectiveness Consideration in the Use of Computer-Assisted Instruction in Economics." Journal of Economic Education, 15:4, pp. 309-323.
- Denham, Carolyn and Lieberman, Ann, eds. (1980). Time to Learn. Washington, D.C.: U.S. Department of Education.
- Diamond, J. J. (1969). A Report on Project GROW: Philadelphia's Experimental Program in Computer Assisted Instruction. Philadelphia: Philadelphia School District.
- diSessa, Andrea A. (1986). "The Third Revolution in Computers and Education." Berkeley, CA: Graduate School of Education, University of California.
- Dunn, A. (1974). Computer-Assisted Instruction Program. A Three Year Report Covering July 1, 1971 Through June 30, 1974. Rockville, MD: Montgomery County Public Schools.
- Fisher, Glenn. "Where CAI is Effective: A Summary of the Research." Electronic Learning, 3:3, pp. 82-83.
- Fletcher, J. D., and Orlansky, Jesse (1986). Cost-Effectiveness of CBI in Defense Training. A paper prepared for the American Educational Research Association Annual Meeting.
- Forsund, Finn, R.; Lovell, C.A. Knox; and Schmidt, Peter (1980). "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement." Journal of Econometrics 13, pp. 5-25.
- Garms, Walter I. (1971). "A Benefit-Cost Analysis of the Upward Bound Program." Journal of Human Resources, 6:2, pp. 206-220 (Spring).
- Glass, Gene V.; McGaw, B.; and Smith, M. L. (1981). Meta-Analysis in Social Research. Beverly Hills, CA: Sage Publications.

- Gramlich, Edward M. (1981). Benefit-Cost Analysis of Government Programs. Englewood Cliffs, NJ: Prentice-Hall.
- Gramlich, Edward M. and Koshel, Patricia P. (1975). Educational Performance Contracting. Washington, D.C.: Brookings.
- Hanushek, Eric A. (1968). The Education of Negroes and Whites. Cambridge, MA: Massachusetts Institute of Technology, doctoral dissertation in Economics.
- Hanushek, Eric A. (1981). "Throwing Money at Schools." Journal of Policy Analysis and Management, 1, pp. 19-42.
- Hanushek, Eric A. (1986). "The Economics of Schooling." Journal of Economic Literature, 24:3, pp. 1141-1177.
- Hatry, Harry P. and Greiner, John M. (1985). "Issues and Case Studies in Teacher Incentive Plans." Washington, D.C.: The Urban Institute.
- Haveman, R. H., and Wolfe, B. L. (1984). "Education, Productivity, and Well-Being: On Defining and Measuring the Economic Characteristics of Schooling." In Dean, E. (ed.): Education and Economic Productivity. Cambridge, MA: Ballinger.
- Hawley, David E.; Fletcher, J. D.; and Piele, Philip K. (1986). Costs, Effects, and Utility of Microcomputer-Assisted Instruction. Eugene, OR: College of Education, University of Oregon.
- Hedges, Larry V. and Olkin, Ingram (1985). Statistical Methods for Meta-Analysis. Orlando, FL: Academic Press.
- Hu, Teh-Wei; Lee, Maw Lin; and Stromdorfer, Ernst W. (1971). "Economic Returns to Vocational and Comprehensive High School Graduates." Journal of Human Resources, 6:1, pp. 23-50 (Winter).
- Jamison, Dean T.; Fletcher, J. Dexter; Suppes, Patrick; and Atkinson, Richard C. (1976). "Cost and Performance of Computer-Assisted Instruction for Education of Disadvantaged Children." In Froomkin, Joseph T., Dean T. Jamison, and Roy Radner (eds.): Education as an Industry. Cambridge, MA: Ballinger.
- Jamison, D. T. and Lau, L. J. (1982). Farmer Education and Farm Efficiency. Baltimore, MD: Johns Hopkins University Press.
- Johnson, Roger T.; Johnson, David W.; and Stanne, Mary Beth (1986). "Comparison of Computer-Assisted Cooperative, Competitive, and Individualistic Learning." American Educational Research Journal, 23:3, pp. 382-392 (Fall).
- Kulik, James A. (1985). "Consistencies in Findings on Computer-Based Education." Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL (April).
- Kulik, James A.; Bangert-Drowns, Robert L.; and Williams, George W. (1983). "Effects of Computer-Based Teaching on Secondary School Students." Journal of Educational Psychology, 75, pp. 19-26.

- Kulik, Chen-Lin C. and Kulik, James A. (1985). Effectiveness of Computer-Based Education of Colleges. Ann Arbor, MI: The University of Michigan, Center for Research on Learning and Teaching.
- Kulik, James A.; Kulik, Chen-Lin C.; and Bangert-Drowns, Robert (1985). "Effectiveness of Computer-Based Education in Elementary Schools." Computers in Human Behavior, Vol. 1, pp. 59-74.
- Lau, Lawrence J. (1979). "Educational Production Functions." In Windham, Douglas M. (ed.): Economic Dimensions of Education. Washington, D.C.: National Academy of Education.
- Levin, Henry M. (1970). "A Cost-Effectiveness Analysis of Teacher Selection." Journal of Human Resources, 5:24-33.
- Levin, Henry M. (1974). "Measuring Efficiency in Educational Production." Public Finance Quarterly, 2:1, pp. 3-24.
- Levin, Henry M. (1976). "Concepts of Economic Efficiency and Educational Production." In Froomkin, Joseph T., Dean T. Jamison, and Roy Radner (eds.): Education as an Industry. Cambridge, MA: Ballinger.
- Levin, Henry M. (1983). Cost-Effectiveness: A Primer. Beverly Hills, CA: Sage.
- Levin, Henry M.; Glass, Gene V.; and Meister, Gail R. (1984). Cost-Effectiveness of Four Educational Interventions. Stanford, CA: Institute for Research on Educational Finance and Governance, Project Report #84-A11.
- Levin, Henry M.; Glass, Gene V.; and Meister, Gail R. (1986). "The Political Arithmetic of Cost-Effectiveness Analysis." Phi Delta Kappan, 68:1, pp. 69-72 (September).
- Levin, Henry M. and Meister, Gail R. (1986). "Is CAI Cost-Effective?" Phi Delta Kappan, 67:10, pp. 745-749 (June).
- Levin, Henry M. and Meister, Gail R. (1986). Educational Technology and Computers: Promises, Promises, Always Promises. IFG Project Report No. 85-A13. Stanford, CA: School of Education, Stanford University.
- Levin, Henry M. and Woo, Louis (1981). "An Evaluation of the Costs of Computer-Assisted Instruction." Economics of Education Review, 1:1-25 (Winter).
- Linn, M. C. and Stein, J. S. (1984). Promise and Reality in Computer Education. Berkeley: University of California, Lawrence Hall of Science.
- Linn, Marcia C.; Rohwer, William D.; and Thomas, John (1986). "Annual Report: Autonomous Classroom Computer Environments for Learning (ACCEL)." Berkeley, CA: Lawrence Hall of Science, University of California at Berkeley.
- Mallar, Charles D. and others (1982). Evaluation of the Economic Impact of the Job Corps Program. Third Follow-Up Report. Princeton, NJ: Mathematica Policy Research.
- Manski, Charles F. (1985). "Academic Ability, Earnings, and the Decision to Become a Teacher: Evidence from the National Longitudinal Study of the High School Class of

- 1972." Cambridge, MA: National Bureau of Economic Research, Working Paper No. 1539.
- McDermott, Cecil W. (1985). Affecting Basic Skills Achievement through Technology. Little Rock: Arkansas State Department of Education.
- Metric Associates (1981). Evaluation of the Computer-Assisted Title I Project, 1980-81. Chelmsford, MA: Metrics Associates.
- Mincer, J. (1979). "Human Capital and Earnings." In Windham, D. M. (ed.): Economic Dimensions of Education. Washington, D.C.: National Academy of Education.
- Murnane, Richard J. and Nelson, Richard R. (1984). "Production and Innovation When Techniques are Tacit: The Case of Education." Journal of Economic Behavior and Organization, 5, pp. 353-373.
- Nelson, Richard R. (ed.) (1982). Government and Technical Change: A Cross-Industry Analysis. New York: Pergamon.
- Nelson, Richard R. and Langlois, Richard N. (1983). "Industrial Innovation Policy: Lessons from American History." Science 219, pp. 814-818.
- Nicholson, Walter (1983). Intermediate Microeconomics and Its Application (third edition). Chicago, IL: Dryden Press.
- Niemiec, Richard P.; Blackwell, Madeline C.; and Walberg, Herbert J. (1986). "CAI Can Be Doubly Effective." Phi Delta Kappan, 67:10, pp. 750-751 (June).
- Niemiec, Richard P.; Samson, Gordon; Weinstein, Thomas; and Walberg, Herbert J. (in press). "The Effects of Computer-Based Instruction in Elementary Schools: A Quantitative Synthesis." AEDS Journal.
- Niemiec, Richard and Walberg, Herbert J. (1985). "Computers and Achievement in the Elementary Schools." Journal of Educational Computing Research, 1, pp. 435-440.
- Niemiec, Richard and Walberg, Herbert J. (1987). "Comparative Effects of Computer-Assisted Instruction: A Synthesis of Reviews." Journal of Educational Computing Research 3:1, pp. 19-37.
- Orlansky, Jesse (1985). The Cost-Effectiveness of Military Training. A paper prepared for the Proceedings of the Symposium on the Military Value and Cost-Effectiveness of Training (DS/A/DR (85) 167). Brussels, Belgium: NATO Headquarters, Defense Research Group on the Defence Applications of Operational Research, January 1985.
- Pogrow, Stanley (1985). "Helping Students Become Thinkers." Electronic Learning, 4:7 (April).
- Pogrow, Stanley (1986a). Policy Recommendations for Developing Appropriate Uses of Technology in Schools. A paper prepared for the National Governor's Association. University of Arizona, School of Education.

- Pogrow, Stanley (1986b). Pedagogical and Curricular Techniques for Using Computers to Develop Cognitive and Social Skills: An Overview of the Techniques Used in the HOTS Program. Tucson, AZ: Thinking With Computers, Inc.
- Psacharopoulos, George (1985). "Returns to Education: A Further International Update and Implications." Journal of Human Resources, 20:4, pp. 584-604.
- Psacharopoulos, George and Loxley, William (1985). Diversified Secondary Education and Development. Baltimore, MD: Johns Hopkins University Press.
- Ragosta, Marjorie; Holland, Paul W.; and Jamison, Dean T. (1982). Computer-Assisted Instruction and Compensatory Education: The ETS/LAUSD Study. Final Report. Princeton, NJ: Educational Testing Service, Project Report #19 (April).
- Rao, V. R. (1977). "Conjoint Measurement in Marketing Analysis." In J. N. Sheth (ed.), Multivariate Methods for Market and Survey Research. Chicago, IL: American Marketing Association.
- Ribich, Thomas I. (1968). Education and Poverty. Washington, D.C.: Brookings.
- Roblyer, M. D. (1985). Measuring the Impact of Computers in Instruction. Washington, D.C.: Association for Educational Data Systems.
- Schoenfeld, Alan (1986a). "Mathematics, Technology, and Higher-Order Thinking Skills in the Near and Not-So-Near Future." Berkeley, CA: Graduate School of Education, University of California (Draft, September).
- Schoenfeld, Alan (1986b). "Use of Computers in Mathematics Instruction." Berkeley, CA: Graduate School of Education, University of California (Draft, October).
- Sexton, Thomas R.; Silkman, Richard H.; and Hogan Andrew J. (1986). "Data Envelopment Analysis: Critique and Extensions." In Silkman, Richard H. (ed.): Measuring Efficiency: An Assessment of Data Envelopment Analysis. San Francisco, CA: Jossey-Bass Inc.
- Shavelson, Richard J.; Winkler, Cathleen Stasz; Feibel, Werner; Robyn, Abby E.; and Shaha, Steven (1984). "Successful Teachers' Patterns of Microcomputer-Based Mathematics and Science Instruction." Santa Monica, CA: Rand Corporation (March).
- Sheingold, Karen; Kane, Janet H.; and Endreweit, Mari E. (1983). "Microcomputer Use in Schools: Developing a Research Agenda." Harvard Educational Review, 53:4, pp. 412-432 (November).
- Sheingold, K. et al. (1981). Study of Issues Related to Implementation of Computer Activity in Schools. Washington, D.C.: National Institute of Education (July).
- Smith, Adam (1937). The Wealth of Nations. New York: Modern Library.
- Sobol, T. and Taylor, R. (1980). "The Scarsdale Project: Integrating Computing into the K-12 Curriculum." Proceedings of the National Education Computing Conference, Norfolk, VA, June 23-25, 1980.

- Spence, M. (1973). "Job Market Signaling." Quarterly Journal of Economics, 87, pp. 355-374 (August).
- Stern, David (1986). "Compensation for Teachers." Review of Research in Education, 13.
- Thompson, Mark (1980). Benefit-Cost Analysis for Program Evaluation. Beverly Hills, CA: Sage Publications.
- U.S. Office of Education (1970). Do Teachers Make a Difference? Washington, D.C.: U.S. Government Printing Office.
- Weisbrod, Burton A. (1965). "Preventing High School Dropouts." In Dorfman, Robert (ed.): Measuring Benefits of Government Investments. Washington, D.C.: Brookings.
- Williamson, Oliver E. (1975). Markets and Hierarchies: Analysis and Antitrust Implications. New York: The Free Press.
- Winkler, John D.; Stasz, Cathleen; and Shavelson, Richard (1986). Administrative Policies for Increasing the Use of Microcomputers in Instruction. Santa Monica, CA: Rand Corporation.